



Estimation statistique de la dispersion de la signature infrarouge multispectrale d'un aéronef faiblement résolu

S. Lefebvre – 6 mai 2009

sidonie.lefebvre@onera.fr



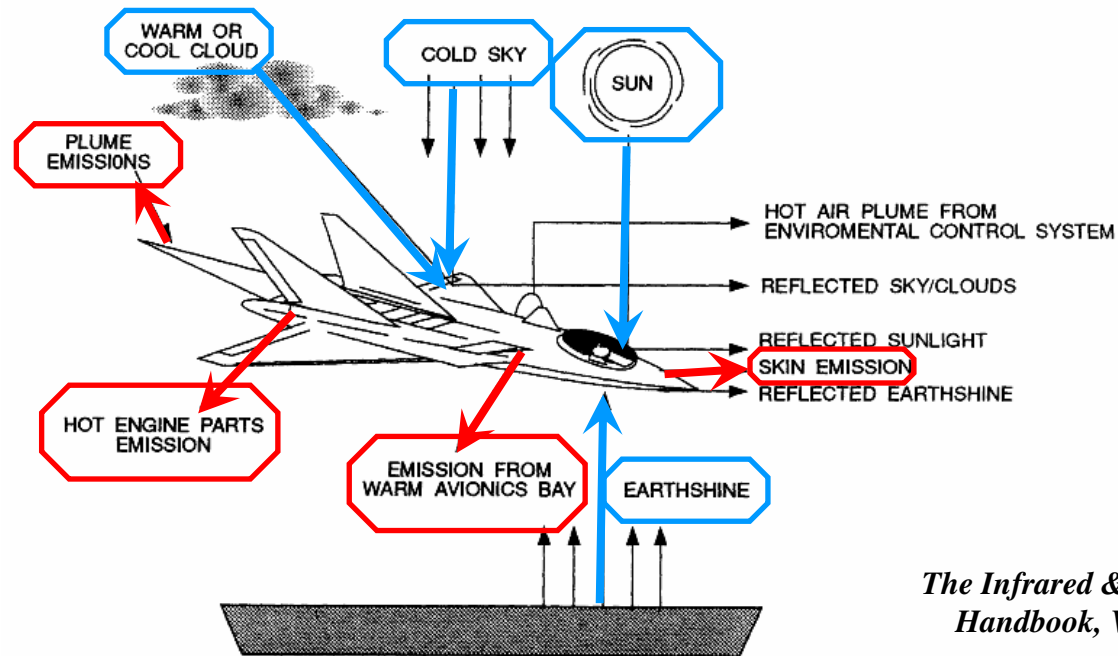
return on innovation



DOTA

Contexte

- Dimensionnement d'un système de veille infra rouge
But : bonne probabilité de détection - taux fausses alarmes faible
- Code CRIRA : calcul SIR aéronef en fonction **propriétés aéronef**
conditions météorologiques
conditions de présentation



The Infrared & Electro-Optical Systems Handbook, Vol 7, Countermeasure Systems

Incertitudes sur données entrée

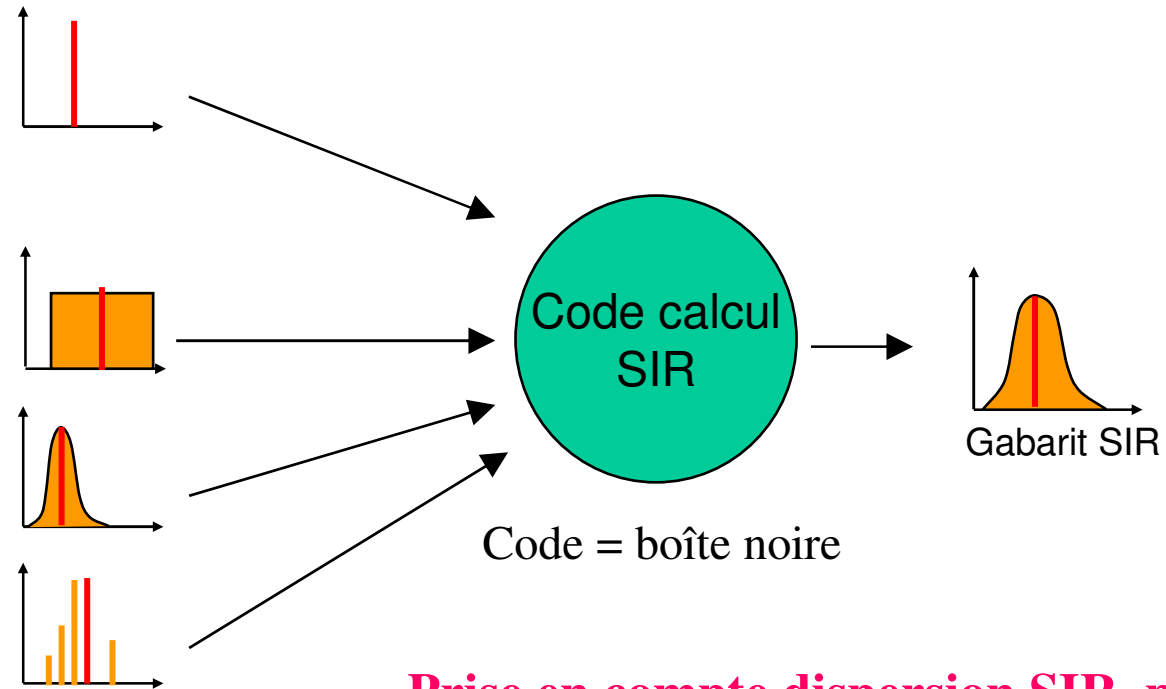
Incertitudes sur les données d'entrée

- Différents types de données :

Données fixes ou paramètres définis par le scénario

Données incertaines :

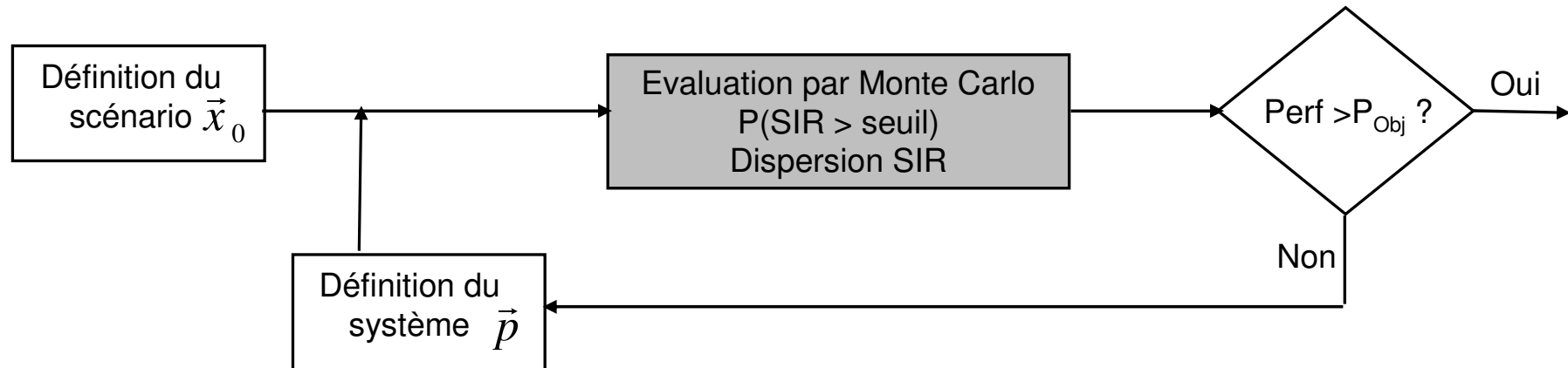
- bornes min-max (peintures, moteurs, orientation aéronef...)
- statistiques (météo...)
- qualitatives (type d'avion, saison, vol au dessus de terre-mer,...)



- Corrélations

Prise en compte dispersion SIR pour dimensionnement système de veille
SIR scalaire : éclaircissement différentiel en entrée de pupille capteur – intégré sur bande large

Démarche



- Identification variables entrée pour scénario : plage variations – couplages – densité probabilité
- Analyse de sensibilité avec plans expériences : classement et **sélection données** influentes SIR
- Dispersion SIR et calcul de P (SIR > seuil) par Quasi Monte Carlo
- Modèle réduit (régression, réseau neurones ...)

Analyse de sensibilité

5000 calculs max. – 28 facteurs – interactions **plan d'expériences factoriel fractionnaire**

Principe : - peu de niveaux pour chaque variable d'entrée – en général 2

- pas de prise en compte des couplages
- passage coordonnées centrées réduites : -1 et +1
- **choix modèle** sous jacent fonction degré précision – importance interactions

$$Y = Cste + \sum_i c_i X_i + \sum_{i < j} c_{ij} X_i X_j + \epsilon$$

Réponse **Effets principaux** **Interactions de degré 2**

- **choix plan**

- réalisation calculs suivant matrice d'expériences obtenue :

$$X_j^i \quad i = 1..n_{\text{fact}} \quad j = 1..n_{\text{calc}} \quad \mathbf{28 \text{ facteurs} - 1024 \text{ exp.}}$$

- exploitation : **calcul effets – stepwise**

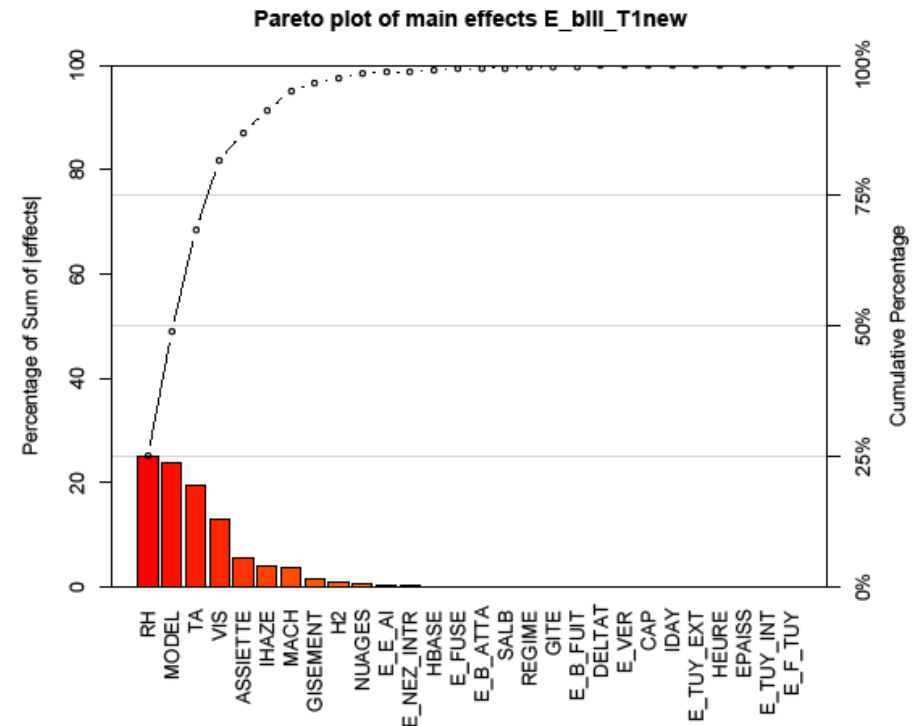
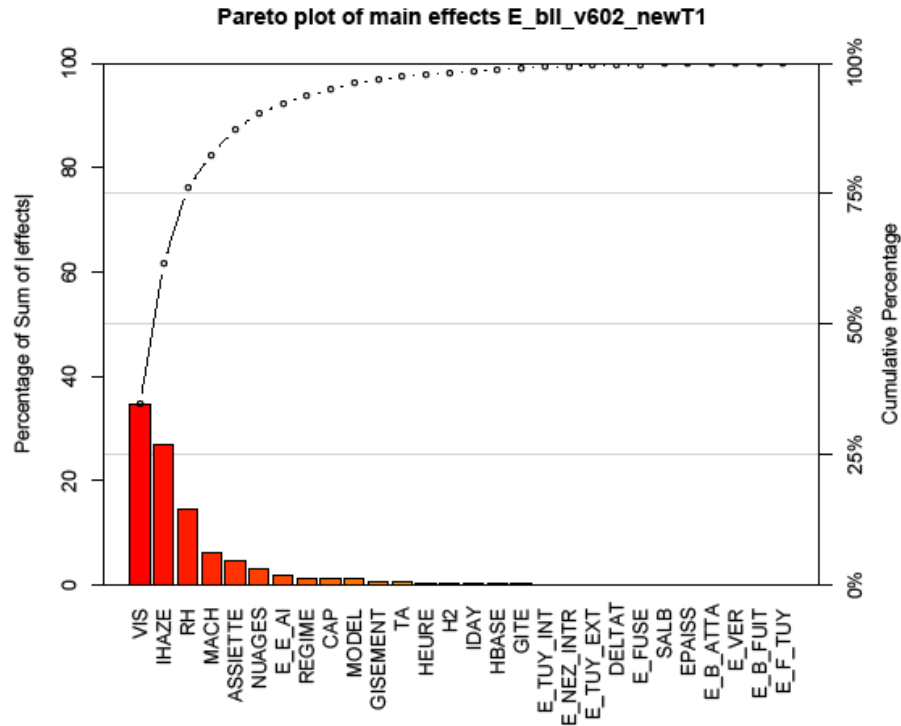
Application : attaque air-sol type

Attaque air-sol de jour, en France, à basse altitude

Diagrammes de Pareto : classement variables sur effets principaux

B II

B III



6 liées conditions météo – 5 aux conditions de vol –
1 aux caractéristiques de l'aéronef

6 liées conditions météo – 4 aux conditions de vol –
2 aux caractéristiques de l'aéronef

12 Facteurs qui contribuent le plus à la variabilité de la SIR



Calcul de P (SIR > seuil) par Quasi Monte Carlo

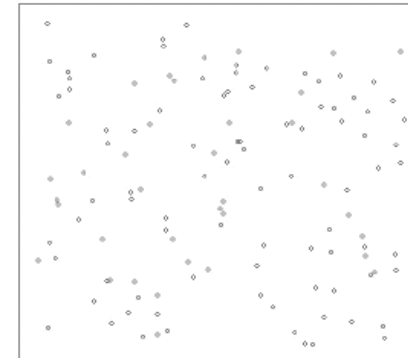
$$P(SIR(X_1, \dots, X_n) \geq \alpha) = \int_{\mathbb{R}^n} \mathbf{1}_{SIR(X_1, \dots, X_n) \geq \alpha}(\vec{t}) p_{X_1, \dots, X_n}(\vec{t}) d\vec{t}$$

- **Monte Carlo:**

$$P(SIR > \alpha) \approx \frac{1}{N} \sum_{i=1}^N I(SIR(u_i) > \alpha)$$

u_i indépendants – loi uniforme erreur $O(1/\sqrt{N})$

$u_i = (X_i^1, X_i^2, \dots, X_i^n)$ $n =$ nombre données entrée



- **Alternative : Quasi Monte Carlo**

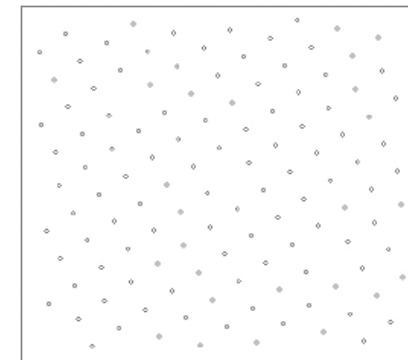
u_i indépendants – termes **suite déterministe à discrédance faible**

discrédance = mesure écart / distribution uniforme

Théorème de **Koksma-Hlawka** :

$$\left| \frac{1}{N} \sum_{j=1}^N f(\xi_j) - \int_{[0,1]^n} f(u) du \right| \leq V(f) D_N^*(\Gamma)$$

Discrédance faible D_N^* : erreur en $O\left(\frac{(\log N)^n}{N}\right)$



cv + rapide que MC si $n \ll$ « pas trop grand »

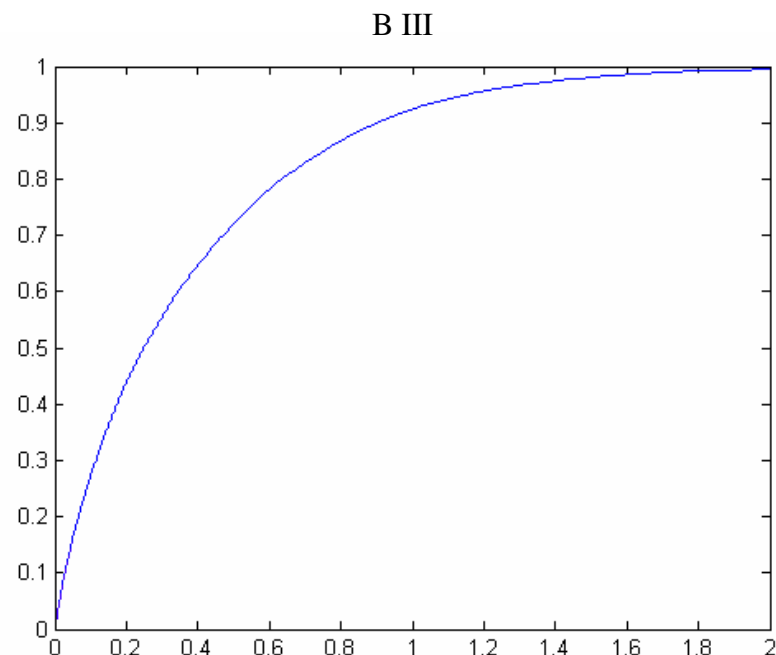
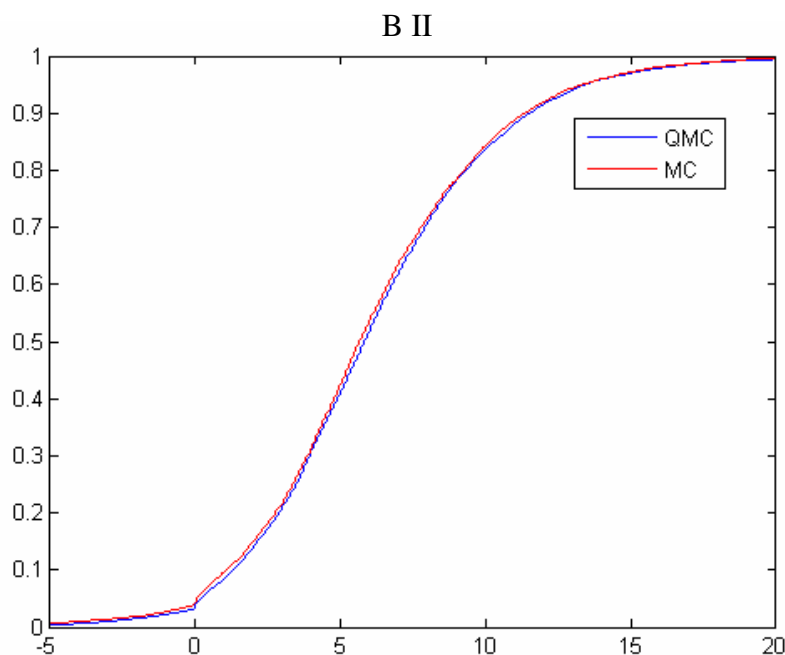
Pb : évaluation de $D^*(N, \xi) \Rightarrow$ majorations inutilisables

En pratique cv plus rapide que MC \Rightarrow dimension effective ?

Résultats : dispersion de la SIR

10240 calculs

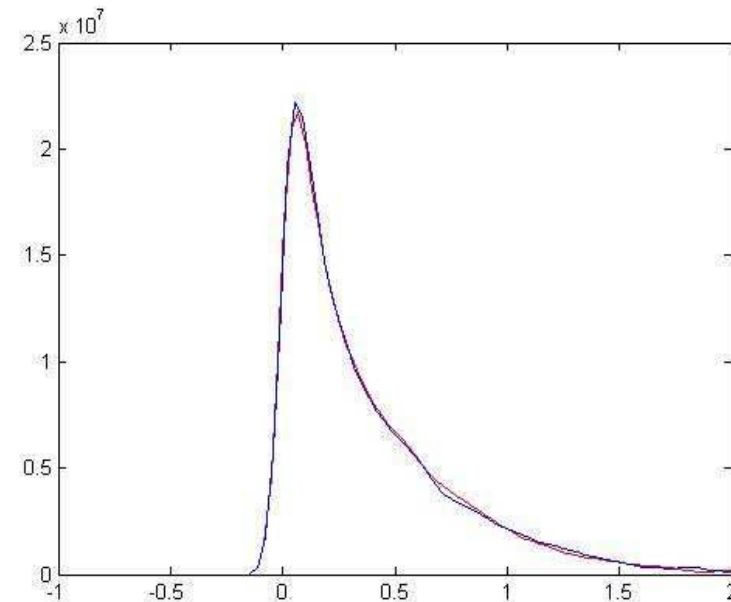
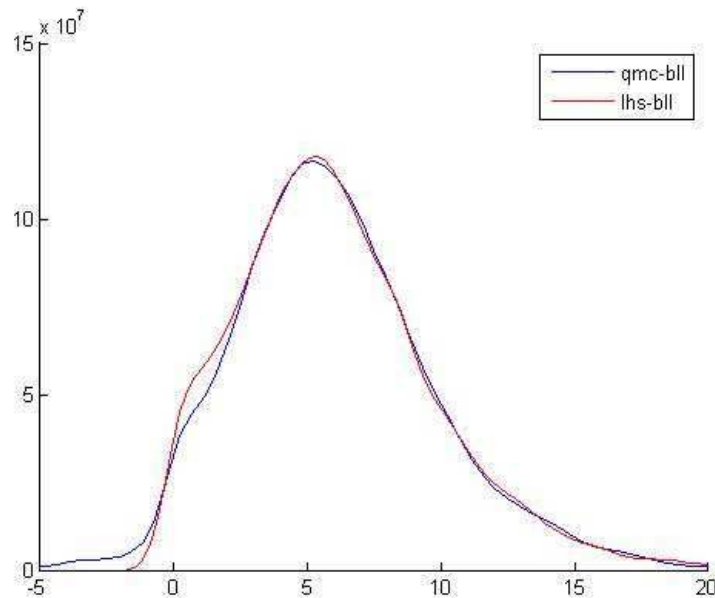
- 8 Variables liées conditions météo : bootstrap dans bdd site *meteo.infospace.ru*
- Autres variables:
 - 14 non influentes : constantes
 - 6 influentes : suite à discrédance faible de Faure avec scrambling



Fonction de répartition empirique => estimation de $P(\text{SIR} > \text{seuil})$



Estimation de la densité de probabilité de la SIR



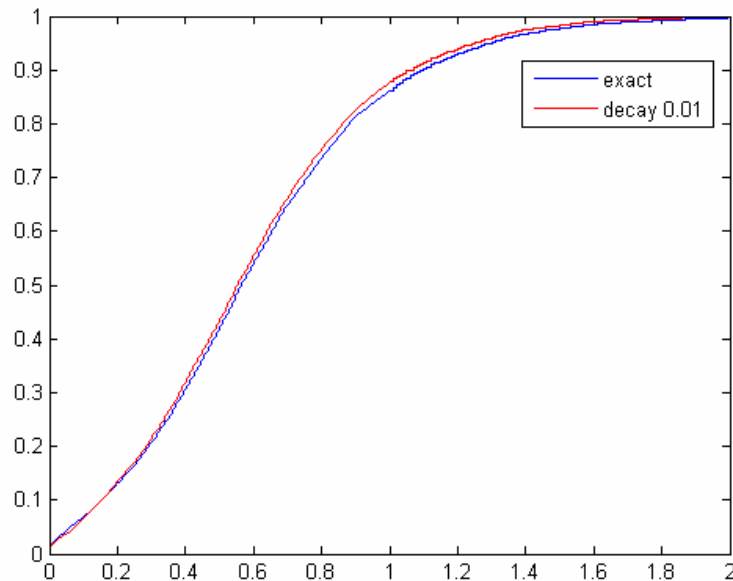
Densité de probabilité estimée en bande II
sur les valeurs de SIR non nulles
parmi les 10240 obtenues en échantillonnant
par **Quasi Monte Carlo et bootstrap**
et par **hypercubes latins et bootstrap**

Densité de probabilité estimée en bande III
sur les valeurs de SIR non nulles
parmi les 10240 obtenues en échantillonnant
par **Quasi Monte Carlo et bootstrap**
et par **Monte Carlo et bootstrap**

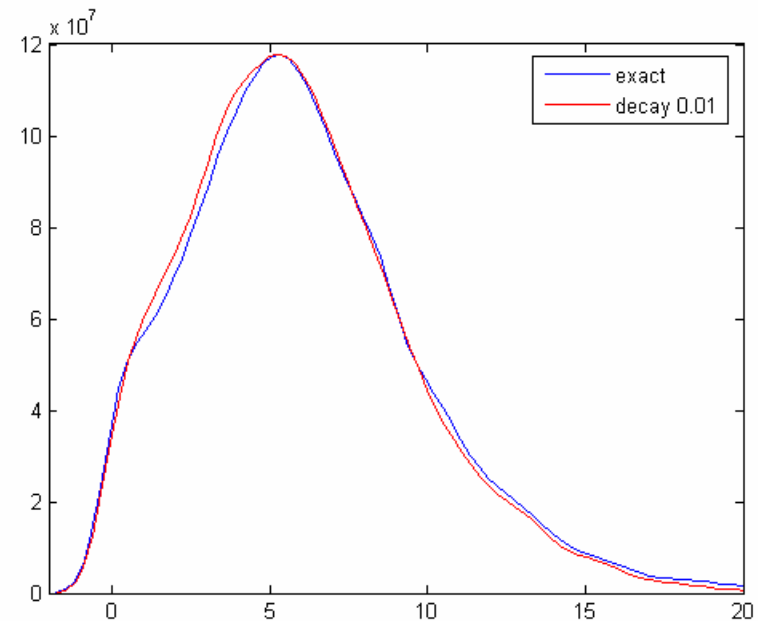
Métamodèle type réseau de neurones – Bande II

11 entrées perceptron multicouches 1 couche cachée – 7 neurones cachés

Apprentissage : échantillon de 4000 points tirés aléatoirement dans 10240 QMC + bootstrap Test: 10240 MC + bootstrap



Fonction de répartition empirique



Densité de probabilité empirique

Très bon accord entre cdf et pdf réelles et prédites

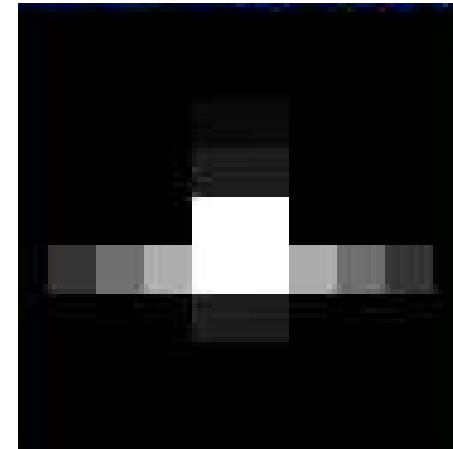
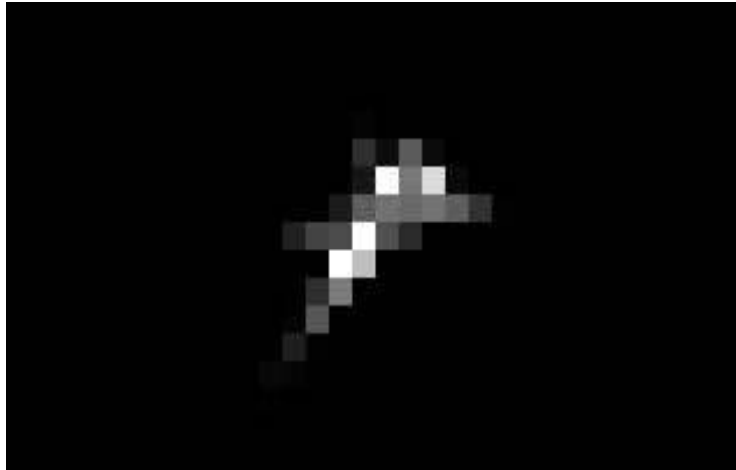
Le métamodèle donne très bonne estimation des probabilités de détection :

différence < 0.5 %

Extension : avion faiblement résolu – SIR spectrale

Sortie = image 7x7 - 10x10 pixels environ et non scalaire
Fond = bruit blanc ou texturé

un spectre / pixel



Questions : - comment réaliser une analyse de sensibilité du code avec sortie vectorielle spatialement et spectralement ?

- comment décrire la dispersion de la SIR ?

2 sources de dispersion : spectrale + spatiale

traduction de la dispersion des images en terme de reconnaissance

- quel critère associer à la dispersion de la SIR :

Critère de Bayes : rapport de vraisemblance ?

Estimation à l'aide d'un métamodèle ?

Approche par étapes :

1. Cas scalaire – intégré
2. Détection cas faiblement résolu – intégré – fond uniforme
3. Identification cas faiblement résolu – intégré – fond uniforme :
3 types cibles – 1 FA
pour un type : dispersion SIR
4. Impact texture fond sur probabilité de détection
5. Prise en compte de l'aspect spectral : thèse 2010 couplée avec thèse expérimentale
But : aider conception de capteurs multispectraux

Pb : comment choisir de façon optimale un petit nombre de bandes spectrales (<10) pour une application donnée (détection, identification...) compte tenu de la dispersion de la SIR ?



Détecteur d'anomalie multispectral

Détecteur RX de Reed et Yu :

$$\begin{cases} H_0 : \vec{x} = N(\mu, \Gamma) & s = [s_1, \dots, s_J] \text{ inconnu, représentant la cible, contenue dans un pixel} \\ H_1 : \vec{x} = N(s, \Gamma) & \vec{x} \text{ le signal enregistré dans les différentes bandes spectrales sur un pixel donné} \end{cases}$$

Γ matrice de covariance spectrale du fond $B - J$ nombre de bandes

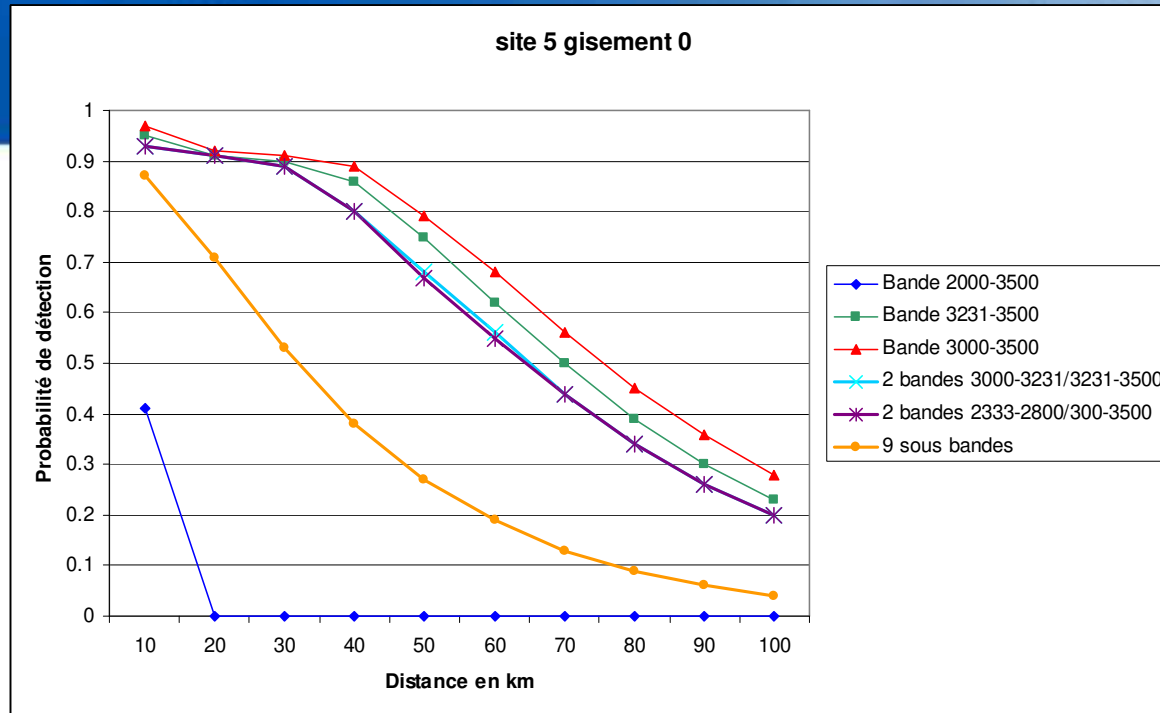
La détection d'anomalie consiste à calculer la probabilité que le signal \vec{x} ne provienne pas du fond

Pour cela, on utilise le détecteur RX, dont la forme est celle de la **distance de Mahalanobis** :

$$D_{RX}(\vec{x}|B) = \left[(\vec{x} - \mu)^T \Gamma^{-1} (\vec{x} - \mu) > t \right]$$

avec t le seuil au-delà duquel le pixel sera considéré comme une anomalie, cad différent du fond B
 t est le paramètre qui sert à contrôler le taux de fausses alarmes





- choix physique : éliminer bandes absorption – bandes SNR faible
- détermination d'un critère : gain en GSNR
 - formule analytique pour 2 ou 3 bandes
- on cherche la solution optimale pour un nombre de bandes fixés, entre 2 et 10
il faut prendre en compte les regroupements de petites bandes
- on compare les solutions pour les différents nombres de bandes
- algo génétique pour trouver combinaison optimale sans tout explorer
- autre critère possible : utilisation stat. 3ème et 4ème ordre
- prise en compte des différentes conditions de présentation :
sur le critère : gain moyen maximal ou moyenne pondérée des gains max. ou gain minimal max....